# ANALYSIS OF BIRTH DATA PROJECT REPORT

## Introduction

The "Analysis of Birth Data" project aims to explore and visualize birth data to gain insights into birth trends and patterns. Understanding birth patterns can provide valuable information for healthcare planning, resource allocation, and identifying potential influences on birth rates.

## Dataset and Data Cleaning

The dataset used in this project, "births.csv," contains information about births, including year, month, day, gender, and the number of births. To ensure data integrity, the following data cleaning steps were performed:

*Replacing Null Values:* Null values in the "births" column were replaced with the average of the existing values to maintain data consistency.

*Removing Outliers:* Values outside the normal range for day (1 to 31) and month (1 to 12) were removed from the dataset to eliminate any erroneous entries.

*Filtering Gender:* Rows with gender values other than "F" (Female) or "M" (Male) were excluded from the dataset to focus solely on valid gender data.

## Visualization 1: Number of Births by Year and Gender

The first visualization involves plotting the number of births by year and gender. A bar plot was created to represent the data, showing separate bars for Female and Male births for each year. The y-axis represents the total number of births, and the x-axis represents the years.

The bar plot provides insights into any trends or patterns in birth rates over the years, comparing the number of births between genders.

**Visualization 2: Distribution of Births by Month and Gender**

The second visualization examines the distribution of births by month and gender. To achieve this, the dataset was grouped by month and gender, and the mean number of births for each group was calculated. A line plot was created to visualize the mean number of births for Female and Male births across each month.

The line plot allows us to identify any seasonality in birth rates throughout the year, as well as differences in the distribution between genders.

**Visualization 3: Number of Births by Day of the Week**

The third visualization focuses on the number of births by the day of the week. The dataset was grouped by day, and the total number of births for each day was calculated. A bar plot was generated to represent the data, with the x-axis representing the days of the week and the y-axis displaying the number of births.

The bar plot helps reveal any patterns or variations in birth rates based on the day of the week.

**Conclusion**

The "Analysis of Birth Data" project successfully explored and visualized birth data, providing valuable insights into birth trends, seasonal patterns, and variations by gender and day of the week. The visualizations offer a clear and concise understanding of the dataset, making it easier to draw meaningful conclusions.

This project can be extended to include additional analyses, such as studying birth rates in specific regions or investigating correlations between birth rates and socio-economic factors. Such insights can

contribute to informed decision-making in healthcare and policy planning.

Thank you for taking the time to explore the Birth Data Analysis project! If you have any questions or feedback, feel free to reach out to me at osuolalefolarin@gmail.com

GitHub Repository: https://github.com/Folarinosuolale/Data-Science-Machine-Learning